How top-down processing enhances comprehension of noise-vocoded speech

How top-down processing enhances comprehension of noise-vocoded speech: Predictions about meaning are more important than predictions about form

Ruth E. Corps[1] & Hugh Rabagliati[1]

[1] School of Philosophy, Psychology, and Language Sciences, University of Edinburgh

Authors' accepted manuscript – in press in *Journal of Memory and Language*

Please address correspondence to:

Ruth Elizabeth Corps

School of Philosophy, Psychology, and Language Sciences

7 George Square

University of Edinburgh

Edinburgh EH8 9JZ

United Kingdom

Ruth.Corps@ed.ac.uk

Word count: 12068 (excluding references & appendix)

Abstract

Listeners quickly learn to understand speech that has been distorted, and this process is enhanced when comprehension is constrained by higher-level knowledge. In three experiments, we investigated whether this knowledge enhances comprehension of distorted speech because it allows listeners to predict (1) the meaning of the distorted utterance, or (2) the lower-level wordforms. Participants listened to question-answer sequences, in which questions were clearly-spoken but answers were noise-vocoded. Comprehension (Experiment 1) and learning (Experiment 2) were enhanced when listeners could use the question to predict the semantics of the distorted answer, but were not enhanced by predictions of answer form. Form predictions enhanced comprehension only when questions and answers were significantly separated by time and intervening linguistic material (Experiment 3). Together, these results suggest that high-level semantic predictions enhance comprehension and learning, with form predictions playing only a minimal role.

Keywords: perceptual learning; noise-vocoding; prediction; speech; dialogue

Introduction

Speech perception is robust and resilient, such that we are able to comprehend utterances across a variety of noisy situations and adverse circumstances. For example, listeners can comprehend speech produced by different talkers at different rates (e.g., Miller & Liberman, 1979) and with different accents (e.g., Clarke & Garrett, 2004). Even when faced with novel acoustic distortions that might make the speech unintelligible at first, listeners can quickly adapt, such that repeated exposure to the distorted speech leads to rapid improvement in comprehension (e.g., Dupoux & Green, 1997). This adaptation is a form of perceptual learning – "relatively long-lasting changes to an organism's perceptual system that improve its ability to respond to its environment and are caused by its environment" (e.g., Goldstone et al., 1998, p. 586).

An ongoing debate for theories of how we understand spoken language has concerned the interaction between higher-level knowledge and lower-level input, and whether listeners immediately use what they know to interpret what they hear (e.g., Magnuson, Mirman, Luthra, Strauss, & Harris, 2018; McClelland & Elman, 1986; Norris, McQueen, & Cutler, 2000). But for the case of how we *learn* to perceive speech, it is generally accepted that higher-level knowledge plays an important role in reorganizing lower-level processing. For example, even for theories that minimize the role of interactivity in speech perception, it is still assumed that high-level knowledge (e.g., of words) influences how people learn to process speech, such as for setting categorical phonetic boundaries and for learning to understand ambiguous fricatives (e.g., McQueen, Cutler, & Norris, 2006; Norris, McQueen, & Cutler, 2003).

There is ample evidence that as listeners process language, they use their high-level knowledge to predict upcoming linguistic information, from the topic of discourse under discussion to the forms of particular words (see Pickering & Gambi, 2018, for a review). A

number of theories, such as predictive coding accounts (e.g., Arnal & Giraud, 2012; Sohoglu & Davis, 2016), assume that the process of perceptual learning is facilitated by prediction. But what types of prediction facilitate learning? In this paper, we address this question by investigating whether high-level knowledge facilitates learning because listeners can predict the semantic content of the distorted words they will hear, or because they can predict the detailed form of these distorted words.

We distinguish between these two possibilities in three experiments that test how people learn to understand noise-vocoded speech, which is an acoustic distortion that smooths large portions of the speech signal's spectral information while preserving temporal cues (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). With the appropriate degree of distortion, vocoded speech can be made approximately 50% intelligible for naïve listeners (Shannon, Fu, & Galvin, 2004), but importantly, the ability to comprehend vocoded speech improves quite quickly with exposure (e.g., Davis et al., 2005), thus making it ideal for investigating the phenomenon of perceptual learning.

In the rest of the Introduction, we review research investigating how high-level knowledge (particularly prediction) aids perceptual learning of noise-vocoded speech. We then describe the current study and formulate our predictions in more detail.

Top-down processing and noise-vocoded speech

A number of studies have shown that noise-vocoded speech is easier to understand and perceived to be clearer if listeners have high-level knowledge of what they are going to hear. For example, Giraud et al. (2004) found that participants could more easily recognize the words in noise-vocoded sentences when they were first been presented with a clear version of the spoken sentence. Similar results were reported by Sohoglu, Peelle, Carlyon,

and Davis (2012; see also Wild, Davis, & Johnsrude, 2012): Participants gave higher clarity ratings to noise-vocoded words when they were presented with matching text prior to hearing the vocoded stimulus. Together, these results suggest that the top-down use of lexical and sentential information in memory can facilitate subsequent processing of that same information when distorted, a phenomenon known as perceptual pop-out.

Importantly, this form of top-down processing not only influences perceptual pop-out and listeners' ability to understand particular tokens of noise-vocoded speech, but also affects perceptual learning and the ability to understand novel vocoded sentences. On each trial of a study by Davis et al. (2005), participants first listened to a noise-vocoded sentence and then transcribed what they had heard. After transcribing this distorted sentence, participants then heard (Experiment 2) or read (Experiment 3) a clear version of the sentence followed by the same distorted version a second time (distorted(D)-clear(C)-distorted(D) condition), or they instead heard the distorted sentence twice before hearing the clear version (DDC condition). The authors found that listeners who knew the identity of the distorted sentence prior to its second presentation (DCD condition) could report more words during the first presentation of subsequent vocoded sentences than participants who heard both versions of the distorted sentence before the clear version (DDC condition). In other words, listeners showed more rapid perceptual learning when they knew the identity of the distorted sentence (and could use information from the clear sentence, in a top-down fashion, to process the distorted sentence) prior to its second presentation.

This learning effect did not occur when participants were trained with non-word sentences (Davis et al., 2005; Experiment 4). However, subsequent work by Hervais-Adelman, Davis, Johnsrude, & Carlyon (2008) found that participants trained with single non-words using a DCD procedure did show comparable perceptual learning to participants trained with words. This discrepancy may be explained by differences in the memorability of

the stimuli in the two studies. Specifically, learning may not have occurred for non-word sentences because participants had difficulty maintaining a string of clear non-words in capacity limited phonological memory (cf. Gathercole, Willis, Baddeley, & Emslie, 1994) and so they could not make comparisons between the clear and distorted versions of the stimulus. When participants were trained with single non-words, however, the phonological representation of the clear form was likely still active in memory when the subsequent distorted version was presented. Thus, top-down facilitation of perceptual learning can occur for non-words if listeners can easily make comparisons between the clear and distorted stimuli.

One open question, however, concerns what mechanisms and information support the top-down facilitation of learning. The most prominent account for how high-level knowledge facilitates learning relies on the prediction of upcoming information, and predictive coding in particular (e.g., Arnal & Giraud, 2012; Sohoglu & Davis, 2016). Predictive coding theories postulate that listeners use their prior knowledge (e.g., from context) to make highly specified moment-to-moment predictions about upcoming events, in a manner that is consistent with a number of other recent theories of language processing, which assume that listeners rapidly use linguistic context to generate predictions about the words that they will hear next (e.g., Christiansen & Chater, 2016). These predictions are immediately compared with incoming linguistic information, and the difference between the two (the prediction error) is carried forward to adjust future processing. For example, listeners presented with a clear version of the stimulus prior to distortion (i.e., in the DCD training condition in Davis et al.'s study) could use this clear representation to precisely predict the form of the distorted input. Any difference between the predicted and actual form yields an error signal, which is used to adjust future predictions, so that they more closely match the incoming speech input.

The best current evidence for the predictive coding account comes from cognitive neuroscience work. For example, Blank and Davis (2016; see also Sohoglu & Davis, 2016; Wild et al., 2012) investigated how manipulations of bottom-up and top-down processing influence the neural responses to distorted speech using fMRI. Participants listened to words (e.g., *sing*) vocoded using either four or twelve bands. These words were preceded by matching written text (e.g., "sing"), partially mismatching written text (e.g., "sit"), or totally mismatching written text (e.g., "doom"). The authors found that presenting matching text and increasing the sensory detail in the auditory stimuli both improved word report scores and reduced BOLD signals in the lateral temporal lobe, a region of the brain that is associated with hearing and comprehending speech (e.g., Davis & Johnsrude, 2003). But multivariate analyses suggested that sensory detail also interacted with the degree of match in the text. In particular, when prior knowledge was uninformative (i.e., mismatching or neutral text), then increases in sensory detail led to an increase in the amount of syllabic information represented in the lateral temporal lobe (quantified using representational similarity analysis; Kriegeskorte & Bandettini, 2008), but when prior knowledge was informative (i.e., matching text), then increases in sensory detail reduced the amount of syllabic information represented in the same area. These results are consistent with a predictive coding account, in which deviations from the predicted input are represented as prediction error. When sensory input mismatches with predictions (i.e., in the mismatching conditions), then prediction error is increased, and so more information about the bottom-up signal is represented. But when sensory input matches with predictions, then the bottom-up input can be explained away, and that information can be discarded.

The studies described so far suggest that top-down processing facilitates perceptual pop-out and perceptual learning by providing participants with the opportunity to generate extremely precise moment-to-moment predictions about the form of what they will hear. But

prediction in these studies has been operationalized using stimulus repetition, such that participants always listened to or saw a clear version of the stimulus before hearing the distorted version, which leaves the characteristics of these predictions somewhat unclear. In particular, it is unclear precisely what information needs to be predicted for enhanced processing and learning to occur. In studies using repetition, a (perhaps implicit) assumption is that by repeating a stimulus, participants should be able to make precise predictions about the form of the distorted words that they will hear, and it is these predictions about form that then facilitate learning by minimizing prediction error.

But a distorted stimulus that is identical to a previously presented clear version is also identical in semantic content, and so top-down effects on perceptual pop-out and learning could also be driven solely by predictions concerning high-level semantic input. For example, if listeners hear or see the clear word *dog* then they could activate the semantic units associated with *dog* (e.g., *four legs, barks*). These semantic predictions need not directly inform perceptual states, like form predictions do in predictive coding accounts, but could instead constrain the processing of ambiguous input and support subsequent learning through feedback connections from semantics to the lexicon (as in a TRACE account of speech perception; e.g., McClelland & Elman, 1986). In fact, when linguistic prediction has been studied outside of the context of stimulus repetition, there has been some controversy about the degree to which listeners and readers tend to generate predictions about the forms of upcoming words (see Pickering & Gambi, 2018, for a review).

Consistent with this argument, studies that showing that the intelligibility of noise-vocoded speech is affected by semantic coherence (in the absence of repetition) suggests that precise sensory predictions are not necessary for perceptual pop-out and learning. For example, Signoret, Johnsrude, Classon, and Rudner (2018; see also Davis, Ford, Kherif, & Johnsrude, 2011) found that clarity ratings were higher for noise-vocoded sentences that were

semantically coherent, and thus constrained the number of potential continuations (e.g., *Her daughter was too young for the disco*), than those that were semantically incoherent, and did not provide any information about the content of the speaker's forthcoming words (e.g., *Her hockey was too tight to walk on cotton*). These findings suggest that participants used the semantic content of the previous words to predict the semantics of forthcoming words, which made them easier to understand in their distorted form. Consistent with previous research showing perceptual pop-out, clarity ratings were also higher when these sentences were preceded by matching rather than mismatching written text. Based on these results, Signoret et al. concluded that both semantic and form-based predictions provide independent aid to perceptual clarity.

However, it is not clear that these two sources of information are truly independent. Participants could use the prior semantic context to generate both content and form-based predictions, for example using the sentence *The boy would like to eat…* to predict that the speaker will refer to an edible object (a semantic content prediction; Altmann & Kamide, 1999) and thus predicting the phonetic features of *cake* (a form prediction). Conversely, they could use the matching text to generate both form and content-based predictions, for example predicting the phonetic features of *cake*, which activates high-level lexical information. Furthermore, Signoret et al. (2018) did not assess perceptual learning, and so it is unclear whether sentence constraint enhances learning in the same way as stimulus repetition (e.g., Davis et al., 2005). Although semantic coherence enhanced perceptual clarity, which means that it was easier for participants to understand the words in distorted sentences for which they had high-level knowledge, it may not make it easier for them to understand novel distorted sentences for which they have no knowledge. In fact, there is also a possibility that apparent perceptual pop-out could partly reflect response bias: When listening to

semantically coherent sentences, it may be easier to guess subsequent words, which may make participants more likely to give higher clarity ratings to distorted sentences.

Overview of experiments

In sum, it is unclear whether high-level knowledge enhances perceptual learning because listeners can use this knowledge to (1) make highly specified predictions about the form of the distorted input, or (2) predict the likely semantic space of possible upcoming words. We discriminated between these two possibilities using three experiments administered online using Prolific Academic. In these experiments, participants listened to question-answer sequences and were asked to type what they thought the answerer said. Using question-answer sequences allowed us to investigate how top-down processes aid perceptual learning without using stimulus repetition.

In all experiments, questions were clearly spoken while answers were noise-vocoded using six channels, which typically produces around 50% intelligibility (e.g., Shannon et al., 2004). Questions were always semantically constraining, and so listeners could use the question to guide their interpretation of the distorted answer. To test whether perception and learning were enhanced by specific form predictions, we manipulated the form constraint of questions so they were either form constraining and predicted a particular answer form (e.g., *What colors are pandas?*; see Table 1), or form unconstraining and did not predict a particular answer form (e.g., *What colors should I paint the wall*?). To test whether listeners used semantic predictions, we also manipulated the semantic consistency of the noise-vocoded answers, so that they were either semantically consistent and made complete sense as a possible answer given the semantic space of the question (e.g., *Black and white*) or semantically inconsistent and made no sense (e.g., *Tom Hanks*).

Table 1. Example materials for the four conditions in Experiments 1 and 2. Note that Experiment 3 uses only stimuli in the semantically inconsistent conditions.

Question Constraint	Question	Answer Consistency	Answer
Form Constraining	What colors are pandas?	Semantically Consistent	Black and white
		Semantically Inconsistent	Tom Hanks
Form Unconstraining	What colors should I paint the wall?	Semantically Consistent	Black and white
		Semantically Inconsistent	Tom Hanks

Experiment 1 assessed whether clear questions could enhance participants' perception of distorted answers, in the same way that stimulus repetition is known to induce perceptual pop-out (e.g., Davis et al., 2005). We refer to the effects we measure as perceptual enhancement, rather than pop-out, in order to account for the fact that the effect is not generated by repetition. Experiment 2 then tested perceptual learning effects, using a manipulation similar to Davis et al.'s DCD condition, to determine whether perceptually enhanced comprehension generalized to novel distorted stimuli. Finally, predictive coding accounts postulate that listeners use in-the-moment predictions to learn, consistent with theories arguing that language processing is "now or never" (Christiansen & Chater, 2016). Experiment 3 investigated the time-course of learning effects to determine whether perceptual enhancement depends on predictions made using the immediate linguistic context.

If perceptual enhancement and perceptual learning effects occur because listeners use high-level knowledge to generate highly specific predictions about form, as would be

expected under a prediction error account, then we expect an interaction between question constraint and answer consistency. When the question is form constraining, listeners can predict the precise form of the answer and can use this prediction to guide their interpretation of the distorted speech. These form predictions are more likely to be accurate when the answer is semantically consistent and makes sense as a response, but inaccurate when the answer is semantically inconsistent. As a result, listeners are likely to correctly report more words in the constraining consistent than the constraining inconsistent condition. In the form unconstraining conditions, however, listeners cannot make highly specified form predictions of the likely answer, and so we expect a smaller difference in the accuracy of word report scores for the semantically consistent and inconsistent answer conditions.

But if top-down effects on perceptual learning are driven by semantic predictions, then we expect listeners to be better at reporting words in distorted answers when these answers are semantically consistent rather than inconsistent, regardless of whether questions are form constraining or unconstraining. In other words, we do not expect an interaction between question constraint and answer consistency. Under this account, participants should use the question (e.g., *What colors should I paint the wall?*) to activate high-level semantic information (e.g., about colors), which should make it easier to integrate the distorted answer when it is semantically consistent (e.g., *Black and white*) than when it is inconsistent (e.g., if participants hear the answer *Tom Hanks*). Given that support for this hypothesis rests on a null interaction, we computed Bayes Factors for all predictors.

Experiment 1

In Experiment 1, participants listened to question-answer sequences, in which the question was clearly spoken while the answer was noise-vocoded, and were asked to type what they thought the answerer said. Importantly, we manipulated the form constraint of

questions, so they either predicted a particular answer form (e.g., *What colors are pandas*?) or did not predict an answer form (e.g., *What colors should I paint the wall*?). These questions were combined with answers that were either semantically consistent, and made sense as a possible response (e.g., *Black and white*), or semantically inconsistent and made no sense (e.g., *Tom Hanks*). Thus, we could test whether high-level knowledge enhances perception only when it allows listeners to predict the specific form of the distorted speech that they will hear, or if general semantic predictions alone are sufficient.

Method

Participants

Eighty native English speakers (21 males; *M*age = 28.56) from Prolific Academic participated in exchange for £1.70. Participants were randomly assigned to one of eight stimulus lists. All participants resided in the United Kingdom and had a minimum 90% satisfactory completion rate from prior assignments. Participants reported no known speaking, reading, or hearing impairments.

Materials

Stimulus Norming

We selected 124 question-answer sequences, 31 for each of the four conditions shown in Table 1, using two norming tasks. First, we selected questions for the two form constraint conditions questions using an online question-answering task, in which 31 further participants from the same population as the main experiment (8 males; *M*age = 20.67) were presented with 62 questions and were instructed to "type your answer into the box below each question. If you do not know the answer, then please guess; do not use Google".

We assessed the form constraint of questions by measuring the degree to which they typically elicited similar answers. To do this, we compared each question's reported answers using Latent Semantic Analysis (LSA; Deerwester, Dumais, Furnas, Landauer, & Harsman, 1990) matrix comparisons using the general reading corpus. LSA determines the similarity of words and phrases by calculating the extent to which they occur in the same context, and ranges from 1 (answers are identical and the question thus constrains the answer) to -1 (answers are completely different and the question is unconstraining).

Using these LSA comparisons, we calculated the constraint of each question by averaging over the LSA values for all pairwise comparisons between answers. Questions in the form constraining condition (M = .86, SD = 0.11) had higher LSA scores than those in the form unconstraining condition (M = .33, SD = 0.15, p < .001 via ANOVA; see Table 2), suggesting they tended to elicit similar answers across participants. In other words, the question constrained the form of the answer. Note that we used the same questions in the semantically consistent and inconsistent conditions, and thus question LSA was identical for each level of answer consistency.

Table 2. Means and (standard deviations) of question LSA scores and answer plausibility for the four conditions in Experiments 1 and 2. Note that Experiment 3 uses only the semantically inconsistent items.

Question Constraint	Answer Consistency	Question LSA[a]	Plausibility Rating[b]
Form Constraining	Semantically Consistent	.86 (0.11)	6.57 (0.47)
	Semantically Inconsistent		1.31 (0.26)
Form Unconstraining	Semantically Consistent	.33 (0.15)	6.09 (0.71)
	Semantically Inconsistent		1.68 (0.82)

[a] Average LSA value over all answer comparisons for that particular question

[b] Plausibility ratings were made on a scale of 1-7. 1 indicated that the answer was very implausible, while 7 indicated that the answer was very plausible.

Using responses from the question-answering task, we selected target answers (between two and four words in length) for stimuli in the form constraining consistent condition (e.g., *What colors are pandas? Black and white*). To ensure that participants heard the same vocoded stimuli across the four conditions, we used the same answers in the form unconstraining consistent condition, even though only 10% of these corresponded to an answer that participants actually provided to the unconstraining questions in the norming task (i.e., these answers were very rarely predicted by participants). We generated answers for the two semantically inconsistent conditions by randomly rotating answers from the semantically consistent conditions. Thus, each answer occurred in all four conditions, but was preceded by a different question (see Table 1).

We assessed the semantic consistency of answers using a second online norming task, in which 44 further participants from the same population (11 males; *M*age = 20.02). were

instructed to: “rate the plausibility of each answer, given the preceding context of the question”. Ratings were made on a scale of 1-7, where 1 indicated that the answer was very implausible (i.e., made no sense and not a possible answer to the question asked) and 7 indicated that the answer was very plausible (i.e., made complete sense and was a possible answer to the question). We randomly assigned participants to one of four lists containing question-answer sequences from all four conditions, and created using a Latin Square procedure, so that each answer occurred once per list.

As intended, answers in the semantically consistent conditions had higher plausibility ratings than those in the semantically inconsistent conditions ($p < .001$ via ANOVA; see Table 2). However, there was also a significant interaction between question constraint and answer consistency. In particular, semantically consistent answers were rated as more plausible when preceded by constraining questions compared to unconstraining questions ($p = .02$), while inconsistent answers were rated as less plausible when preceded by constraining questions compared to unconstraining questions ($p = .002$). Thus, plausibility (and as a result, semantic consistency) was not matched across levels of question constraint. Note that this interaction cannot be attributed to collinearity between question LSA and plausibility ratings, since we found no correlation between these two values ($r = .01$, $p = .89$).

To try and overcome the differences in semantic consistency across levels of question constraint, we conducted a second pre-test of answer plausibility using a different set of rotated answers for the inconsistent conditions. However, we still found the same interaction between answer consistency and question constraint, and so we returned to the first set of rotated answers. It is likely that we were unable to balance plausibility ratings across the two levels of form constraint because constraint either made it easier to identify implausibility. When questions are constraining, for example, there is often only one possible answer (e.g., *When is New Year’s Eve? The thirty first of December*), and so all others are considered

semantically inconsistent and thus implausible because they are likely incorrect. When questions are unconstraining (e.g., *What is your favorite film?*), however, there are a variety of possible answers and so it is not immediately clear which answers are inconsistent.

As a result, we do not analyze the experiment as a factorial design, because any interaction between question constraint and answer consistency could be attributed to the differences in answer consistency across levels of question constraint. Instead, we use the continuous values of question constraint (question LSA) and answer consistency (answer plausibility rating). We return to this issue in the Data Analysis and Results sections.

Stimulus Recording and Vocoding

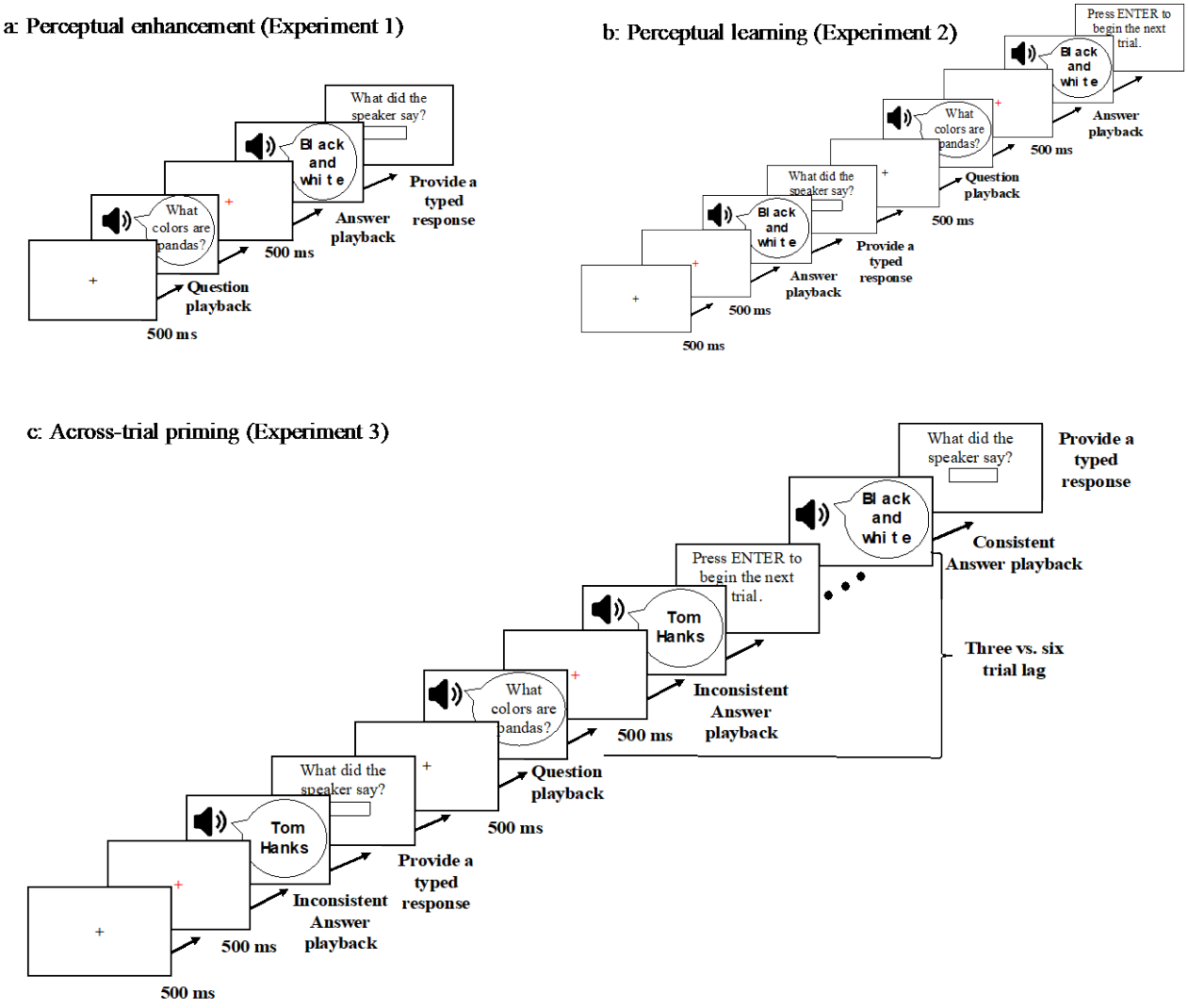
Questions were recorded by a native English female speaker, who was instructed to read the utterance as though "you are asking a question and expecting a response". Answers were recorded separately by a native English male speaker, who was instructed to read the utterances as though "you are answering a question". The amount of sensory detail available in answers was varied using noise-vocoding (Shannon et al., 1995), which divides the speech signal into frequency bands and then applies the amplitude envelope in each frequency band onto corresponding frequency regions of white noise, thus removing spectral information from the signal while still preserving temporal cues. Vocoding was performed with a custom MATLAB (MathWorks) script using six spectral channels logarithmically spaced between 70 and 5000 Hz (Blank & Davis, 2016), and answers were thus unintelligible to naïve listeners.

Procedure

The experiment was administered online. Stimulus presentation was controlled using jsPsych (de Leeuw, 2015) and data was recorded using MySQL (version 5.7). Participants were warned that they would be listening to audio stimuli, and so were encouraged to

complete the experiment in a quiet environment or to use headphones. Before the task, participants were instructed: “First you will hear a female speaker ask a question in a clear voice. You will then hear a male answer this question in a distorted voice. Your task is to listen carefully and type exactly what you think the male speaker said. If you do not know, then please guess”. To make stimulus onset salient, a fixation cross appeared 500ms before question playback (see Figure 1a). The fixation cross then turned red and answer playback began 500ms later. After listening to the answer, participants were prompted to type their response and press a “submit answer” button.

Figure 1. Schematic representation of the structure of Experiments 1, 2, and 3.



Design

Question constraint and answer consistency were manipulated within items but between participants. We manipulated question constraint and answer consistency between participants in order to ensure these data were comparable to our investigation of perceptual adaptation in Experiment 2, which necessarily required a between-subjects design.

Participants were randomly assigned to one of eight stimulus lists, each containing 15 items (one item was discarded to ensure there were an equal number of stimuli in each list). We created eight lists of 15 stimuli, rather than four lists of 31 stimuli, to ensure that answers in the inconsistent conditions appeared in a separate list from their corresponding question, so that they could not be primed by previous exposure (e.g., such that the inconsistent answer *James Bond* would not be primed by an earlier trial such as *Which character is also known as 007? King's Cross*). Participants thus heard only one version of each answer (either consistent or inconsistent) and one version of each question (either constraining or unconstraining), and all items they heard belonged to the same condition. Although we assigned participants to one of four conditions, we used the continuous values of question constraint (question LSA) and answer consistency (answer plausibility rating) when analyzing the results to overcome the differences in answer consistency in the form constraining and unconstraining conditions.

Data Analysis

Although it may seem intuitive to analyze participants' accuracy at reporting words in the heard distorted answer, this analysis is likely to be affected by response bias. In particular, participants may be biased towards reporting an answer that is consistent with the question that they heard, rather than reporting the answer that they actually heard. For example, when the answer was *Black and white* and the question was *What colors are*

pandas? (as in the form constraining semantically consistent condition), then participants who simply reported the expected answer to the question would have 100% accuracy. But if *Black and white* was preceded by *Where does the Queen live?* (as in the form constraining semantically inconsistent conditions), then participants following the same strategy (i.e., reporting the expected answer) would have 0% accuracy. This response bias would lead to an interaction between question constraint and answer consistency, because form constraining questions will bias participants towards a particular answer more than form unconstraining questions (i.e., the bias induced by *What colors are pandas?* is greater than that induced by *What colors should I paint the walls?*). Importantly, this interaction is the same as predicted by a prediction error account, and so analyzing accuracy cannot tell us whether interpretation is affected by form predictions independently of response bias.

To counter this concern, we conducted a signal detection analysis to determine participants' sensitivity to the words they actually heard in the vocoded answers, while controlling for response biases that may have been induced by the preceding question. In this analysis, we assessed whether participants' tendency to report the semantically consistent answer was affected by whether this consistent answer was actually heard. For example, when the question was *What colors are pandas?* or *What colors should I paint the wall*?, then we coded how many words in the participants' answer were components of the associated consistent response (i.e., *Black and white* for both conditions). Reporting *Black and white* having heard *What colors should I paint the wall? Black and white* can be thought of as a Hit, whereas reporting *Black and white* having heard *What colors should I paint the wall? Tom Hanks* can be thought of as a False Alarm.

When coding participants' responses, words with obvious spelling mistakes or typing errors (i.e., from keys around the target letter/word, missing letters, etc.) were simply corrected, but morphological mismatches were not (i.e., reporting *younger* was considered

incorrect if young was expected; see also Davis et al., 2005). Words reported in the wrong order were not scored as matching the consistent answer. Of the 1200 responses, we discarded 14 (1.12%) because participants typed the preceding question rather than the distorted answer.

We could not conduct a standard signal detection analysis because our design was between-subjects, and so we instead formulated our signal detection model as a mixed effects logistic regression model (DeCarlo, 1998; Wright, Horry, & Skagerberg, 2009). In this model, we predicted the proportion of words that participants reported in the expected answer was predicted by Question Constraint (Question LSA), Answer Consistency (Answer Plausibility Rating), Block, and their interaction). Both Question Constraint and Answer Consistency were centered and standardized continuous variables, while Block was a centered numeric predictor (i.e., -1 for trials 1-5, 0 for trials 6-10, and 1 for trials 11-15). For the purpose of data presentation, we plot participants' accuracy at reporting the expected answer, split by factorial levels of Question Constraint, Answer Consistency, and Block.

We fitted models using the *glmer* function of the *lme4* package (version 1.1-14; Bates, Bolker, & Walker, 2015) in RStudio (version 0.99.903). We used the maximal random effects structure justified by our design (Barr, Levy, Scheepers, & Tily, 2013), but correlations among random effects were fixed to zero to aid model convergence (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). The regression thus had the form, in lme4 syntax, cbind(ReportedExpectedWords, UnreportedExpectedWords) ~ Question Constraint * Answer Consistency * Block + (1+Block || Participant) + (1+Question Constraint * Answer Consistency * Block || Item). Note that the intercept reflects overall criterion, which is modulated by Question Constraint (capturing how question constraint affects bias). The effect of Answer Consistency corresponds to overall sensitivity, analogous to a *d'* score (i.e., hits minus false alarms) and thus it captures perceptual enhancement. Most importantly, the

interaction between Answer Consistency and Question Constraint corresponds to whether sensitivity differs depending on the constraint of the question, and is thus important for determining whether form predictions enhance perception. If the effect of Answer Consistency is larger for more constraining questions (i.e., sensitivity is higher when questions are more constraining), then this would indicate that form predictions enhance perception of distorted speech.

For each predictor in the regression, we report coefficient estimates (*b*), standard errors (*SE*), and *p* values. In addition, we computed Bayes factors for all predictors by fitting generalized Bayesian mixed effects models using the *brms* package (version 2.1.0; Bürkner, 2018) with student_t priors (with ten degrees of freedom, a mean of zero, and a standard deviation of one) for all population-level effects. Models were fitted using a binomial distribution, and we ran 4 chains per model, each for 10000 iterations with a burn-in period of 5000 and initial parameter values set to zero. In all instances, we compared the full model to a model excluding the relevant predictor(s). Following Kass and Raftery (1995), we interpret Bayes factors greater than 3.2 (or less than 1/3.2) as substantial evidence for the alternative hypothesis (or for the null), and Bayes factors greater than 10 (or less than 1/10) as strong evidence for the alternative hypothesis (or for the null). Raw data, analysis scripts, and lists of experimental stimuli are available at https://osf.io/y96tq/.

Results

Figure 2 illustrates how responses varied across Question Constraint, Answer Consistency, and Block; the left panel shows the average proportion of words reported in the heard (correct) answer, and the right panel shows average proportion reported the expected answer. As described above, we focus our analysis and interpretation on whether participants' accuracy at reporting the expected answer was affected by Question Constraint

and Answer Consistency, given that accuracy at reporting the heard answer is likely to be affected by response bias.

Figure 2. Observed means of the proportion of words in the heard answer (left panel) and the expected answer (right panel) reported correctly for the four factorial conditions across the three blocks in Experiment 1. Error bars represent ± 1 standard error from the mean.

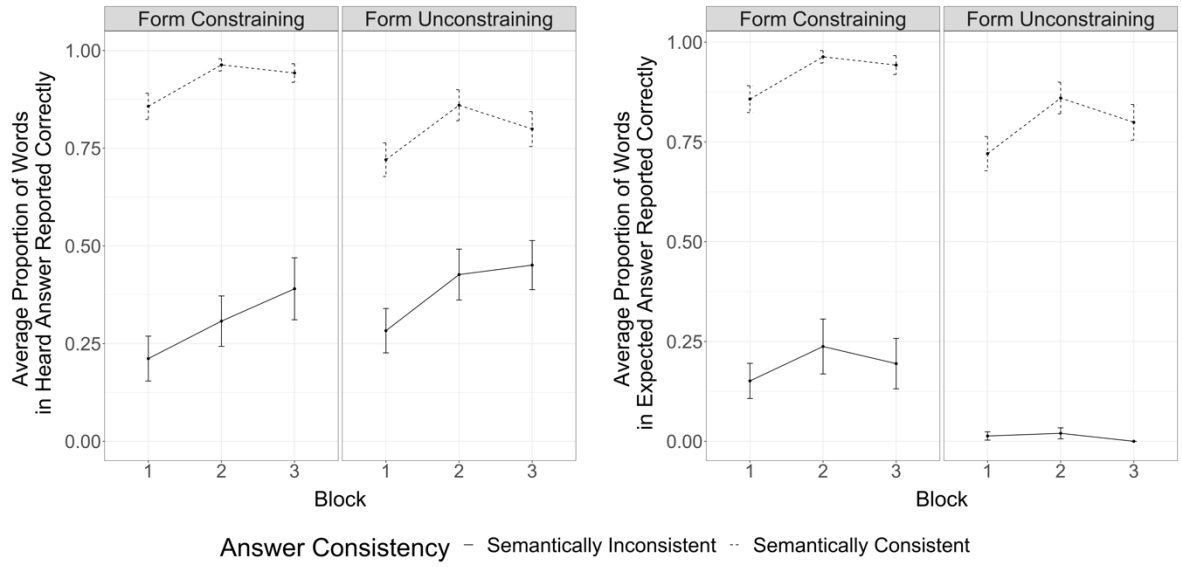


Table 3. Full model output for fixed effects for the analysis of word report scores in Experiment 1.

	Words Reported in Expected Answer			
Predictor	Estimate (*SE*)	*z*	*p* value	Bayes factor
Intercept	-0.66 (0.36)	-1.84	.07	-
Question Constraint	1.19 (0.28)	4.32	< .001	>100
Answer Consistency	4.29 (0.42)	10.11	< .001	>100
Block	0.31 (0.15)	2.06	.04	2.29
Question Constraint * Answer Consistency	-0.18 (0.30)	-0.59	.55	0.19
Question Constraint * Block	0.11 (0.17)	0.67	.51	0.27
Answer Consistency * Block	0.37 (0.20)	1.84	.07	0.50
Question Constraint * Answer Consistency * Block	0.03 (0.22)	0.12	.91	0.42

As the right panel of Figure 2 illustrates, participants' responses were biased by the question that they heard: They were more likely to report the expected answer after they heard a question that was form constraining rather than form unconstraining, no matter whether or not that answer was correct (b = 1.19, SE = 0.28, p < .001). After accounting for response bias, we could then assess participants' sensitivity to the information in the vocoded answer. In this analysis, sensitivity is captured by the effect of Answer Consistency, which was significantly greater than 0 (b = 4.29, SE = 0.42, p < .001; see Table 3).

If form predictions enhance perception of noise-vocoded speech, then we expect an interaction between Question Constraint and Answer Consistency, such that participants should be more sensitive to semantically consistent answers than semantically inconsistent answers, but only when these answers are preceded by form constraining questions. However, we found no evidence for this interaction (b = -0.18, SE = 0.30, p = .55), and the Bayes factor of 0.31 suggested that there was substantial-to-strong evidence for this null effect. Indeed, the sign of the regression coefficient is in fact more consistent with sensitivity being greater when questions are *less* constraining rather than more constraining, which would not be expected if predictions about form enhance processing.

Finally, we found a significant effect of Block, such that participants were more likely to report the expected answer in later than in earlier blocks (b = 0.31, SE = 0.15, p = .04). No further predictors were significant, and the Bayes factors confirmed these null effects (see Table 3).

Discussion

In Experiment 1, we investigated whether high-level knowledge enhances comprehension of distorted speech because it allows listeners to predict the meaning of the distorted utterance, or because it allows listeners to predict the lower-level word forms. We

found clear evidence that being able to predict the semantics of distorted answers enhanced perception: Participants were better at reporting vocoded answers when they were semantically consistent with the question (left panel of Figure 2). However, we did not find any evidence that form predictions helped guide listeners' interpretation of distorted answers. In our signal detection analysis, which estimated participants' sensitivity to information in the distorted speech while accounting for response bias, sensitivity was not enhanced when the preceding question allowed participants to accurately predict the forms of the words that they would hear. In particular, there was no interaction between question constraint and answer consistency, and a Bayes factor analysis suggested substantial-to-strong evidence that the effect was indeed null. This analysis thus suggests that an answer such as *Black and white* was similarly interpretable regardless of whether the question was *What colors are pandas?* (form constraining question, semantically consistent answer) or *What colors should I paint the wall*? (form unconstraining question, semantically consistent answer).

We also found that participants were more likely to report the expected answer in later rather than earlier blocks. In our next study, we investigated this adaptation in more detail, examining whether top-down effects on learning (rather than perception) were affected by precise predictions about form or more general predictions about meaning. We used a similar design to Davis et al.'s (2005) Distorted-Clear-Distorted condition. In particular, participants reported the distorted answer before they heard its corresponding question, thus removing the influence of response bias and allowing us to further determine whether form predictions play a role in learning.

Experiment 2

In Experiment 2, we used a similar design to Davis et al.'s (2005) Distorted-Clear-Distorted condition (described in the Introduction) to investigate whether learning is

enhanced by predictions of form or meaning. Participants first heard a distorted phrase and reported what they heard. They then heard a clear question followed by the same distorted phrase, this time used as an answer to the question (see Figure 1b). As in Experiment 1, we varied the relationship between questions and answers (see Table 1), to determine whether being able to predict the form or semantic content of distorted answers (from the clear question) not only enhances perception of the current answer, but also enhances comprehension of subsequent, unrelated answers.

On the one hand, if high-level knowledge enhances perceptual learning because it allows listeners to predict the form of distorted speech, then we expect an interaction between question constraint and answer consistency. In particular, participants should be better at comprehending novel distorted answers (i.e., on their first presentation) when they have previously heard constraining questions followed by semantically consistent answers compared to when these questions are followed by semantically inconsistent answers, but importantly this difference should be smaller when participants are trained with unconstraining questions. On the other hand, if top-down effects on learning are mainly dependent upon predictions about the semantics of the distorted utterance, then we expect listeners to be more accurate at reporting distorted words when they have previously heard answers that are consistent (rather than inconsistent) responses to questions, regardless of the constraint of that question.

Method

Participants

One hundred and twenty-eight further native English speakers (25 males; *M*age = 20.47) participated. We first recruited 100 participants (19 males; *M*age = 18.44) from the undergraduate student pool at the University of Edinburgh, who participated in exchange for

partial course credit. Using the same procedure as Experiment 1, we recruited the remaining 28 participants (6 males; *M*age = 27.71) from Prolific Academic. We used two different participant samples because some testing occurred outside of semester time, and so we could not recruit all participants in exchange for course credit.

Materials and Procedure

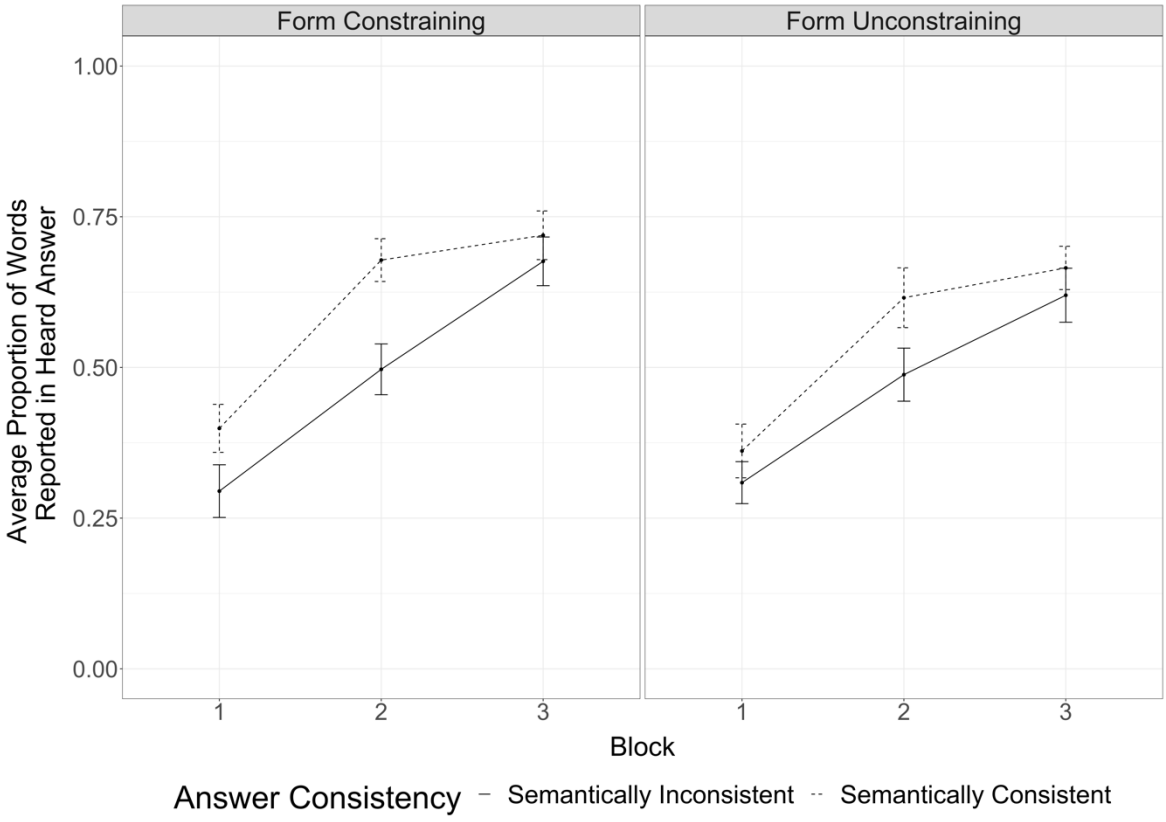
The materials were identical to those used in Experiment 1. Participants were tested online, as in Experiment 1, but using a slightly different procedure (see Figure 1b). Participants were first presented with the distorted answer, followed by the clear question, and then the same distorted answer a second time. Participants saw two fixation crosses before the onset of the first answer: the first (black) to indicate the beginning of the trial, and the second (red) to indicate the onset of the answer. Participants were prompted to type the distorted answer on its first presentation and then listen to the question-answer sequence. In particular, they were told: "First, you will hear a male speaker produce a statement in a distorted voice. Please type the words of that statement in the box provided. You will then hear a female speaker produce the question to that statement in a clear voice. The male speaker will then repeat the distorted statement a second time. You do not need to type this statement a second time; please just listen to the exchange".

Results

Response bias was not an issue in this study design (as participants reported the answer before hearing a question), and so we simply analyzed participants' accuracy in this task, rather than relying on a signal detection analysis. Thus for each trial, we coded the number of words each participant correctly identified in the answer they heard. The regression used a similar multi-level model as in Experiment 1, but

cbind(ReportedExpectedWords, UnreportedExpectedWords) was replaced by cbind(ReportedHeardWords, UnreportedHeardWords). Of the 1920 responses, we discarded six (0.31%) because participants reported the question from the previous trial rather than the answer for the current trial. On average, participants correctly identified 53% (*SD* = 17%) of the words in the distorted answers (see Figure 4 for a breakdown of proportions by factorial conditions and block), and this percentage increased across the three blocks ($b = 0.94$, $SE = 0.10$, $p < .001$), suggesting that participants quickly learned to understand the distorted speech.

Figure 4. Observed means of the proportion of words in the heard answer reported correctly for the four factorial conditions across the three blocks in Experiment 2. Error bars represent ± 1 standard error from the mean.



Interestingly, we found that semantic predictions enhanced perceptual learning: Participants were better at reporting words in novel distorted answers when they had previously heard question-answer sequences in which answers were consistent responses to the preceding question (b = 0.30, SE = 0.11, p = .004; see Table 4). For example, participants were better at understanding the distorted phrase *Black and white* when they heard (on a previous trial) the clear question *Which space ranger starred in Toy Story?* and the distorted answer *Buzz Lightyear* compared to when they heard the same question paired with the answer *The Eiffel Tower.* Additionally, word report scores were not affected by whether participants were affected the constraint of questions participants heard on previous training

trials (b = 0.05, SE = 0.10, p = .61) and the Bayes factor (0.18) suggested strong evidence for this null effect.

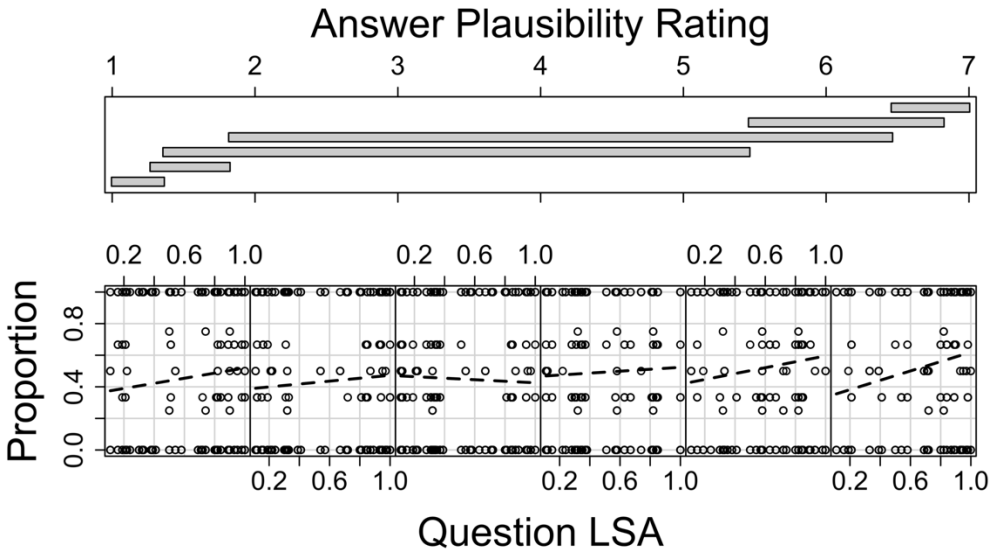
Table 4. Full model output for fixed effects for the analysis of word report scores in Experiment 2.

Predictor	Estimate (SE)	z	p value	Bayes factor
Intercept	0.27 (0.28)	0.98	.33	-
Question Constraint	0.05 (0.10)	0.50	.61	0.18
Answer Consistency	0.30 (0.11)	2.89	.004	8.36
Block	0.94 (0.10)	9.24	< .001	>100
Question Constraint * Answer Consistency	0.02 (0.10)	0.18	.86	0.15
Question Constraint * Block	-0.02 (0.08)	-0.19	.85	0.00
Answer Consistency * Block	0.04 (0.08)	0.52	.60	0.19
Question Constraint * Answer Consistency * Block	-0.17 (0.09)	-1.81	.07	0.98

Importantly, we did not find any evidence that predictions about form played an important role in perceptual learning. In particular, there was no significant interaction between the form constraint of the question and the consistency of the answer (b = 0.02, SE = 0.10, p = .86), such that word report scores were similar for consistent and inconsistent answers, regardless of question constraint. The Bayes Factor of 0.15 indicates strong evidence for this null effect. This interaction is illustrated in Figure 5, which shows that the

effect of Question Constraint did not vary substantially for different levels of Answer Consistency.

Figure 5. The relationship (represented by points and regression lines) between the proportion of words correctly identified and Question LSA at each level of Answer Plausibility Rating in Experiment 2. Note that each facet represents a different level of answer plausibility.



However, we did find a marginally significant three-way interaction between Question Constraint, Answer Consistency, and Block (although note the Bayes Factor of 0.97 for the interaction suggests that the evidence for such an interaction is weak). To follow up this potential interaction, we fitted separate models for each block. These analyses showed that participants were marginally better at identifying words in distorted answers with higher Answer Consistency in Block 2 (b = 0.38, SE = 0.21, p = .07), but not in Blocks 1 (b = 0.29, SE = 0.21, p = .15) or Block 3 (b = 0.30, SE = 0.23, p = .19). Furthermore, participants were marginally better at identifying words in distorted answers when they were preceded by questions that were more rather than less constraining in Block 3 (effect of Question

Constraint; b = 0.44, SE = 0.23, p = 0.06) but not in Block 1 (b = -0.02, SE = 0.23, p = .92) or Block 2 (b = 0.32, SE = 0.21, p = .13). But importantly, and inconsistent with an account in which listeners were learning by using high-level knowledge to make highly specified form predictions of the distorted input, there was no interaction between Answer Consistency and Question Constraint in any of the blocks (Block 1: b = 0.09, SE = 0.20, p = .66; Block 2: b = 0.16, SE = 0.22, p = .48; Block 3: b = 0.24, SE = 0.28, p = .38).

Discussion

In Experiment 2, we investigated how top-down processing enhances perceptual learning of distorted speech by manipulating: (1) the constraint of questions, to determine whether high-level information is informative for learning when it allows listeners to predict the specific form of the speech they will hear, and (2) whether the noise-vocoded answer was a semantically consistent response to the question, to determine whether general semantic predictions alone are sufficient for learning. By instructing participants to report the noise-vocoded answer before they heard its corresponding question, we could assess how experience with question-answer sequences on previous trials helped participants learn to comprehend novel distorted answers that they had never heard before, while also removing the potential for response bias that we observed in Experiment 1.

Participants were better at identifying words in novel noise-vocoded answers when they were trained with question-answer sequences in which the answer was a semantically consistent response to the preceding question, suggesting that high-level predictions about semantic content enhanced learning. However, word report scores were unaffected by question constraint and there was no interaction between question constraint and answer consistency. This lack of interaction suggests that form predictions did not help participants learn to understand distorted speech. This finding is consistent with the signal detection

analysis in Experiment 1, which suggested that (once response bias was accounted for) predictions about form did not enhance perception. Together, these studies suggest that being able to predict the form of distorted speech does not enhance learning and comprehension of that speech, above and beyond any effects of semantic prediction. We suggest that this finding may be inconsistent with versions of a predictive coding hypothesis, in which learning depends upon mismatches between the predicted form of the stimulus and the actual distorted input. We discuss the theoretical implications of these findings in more detail in the General Discussion.

In sum, Experiments 1 and 2 demonstrate that high-level information enhances interpretation of and perceptual learning about distorted speech because it allows listeners to predict the semantic content of a distorted phrase. In contrast, predictions about the forms of words did not enhance interpretation and learning above-and-beyond any effects of semantic predictions. One possibility is that lower-level predictions simply cannot influence interpretation and learning. Experiment 3 addresses this issue.

Experiment 3

Experiments 1 and 2 suggest that high-level knowledge enhances perception and learning because it allows listeners to predict the semantic features associated with the distorted input, which then presumably facilitates integration of novel distorted speech representations into pre-existing higher-level representations. In these experiments, we focused on the effects of high-level knowledge when listeners could predict the distorted input using the immediately surrounding linguistic context (i.e., from presentation of the question immediately before the distorted answer).

In fact, prediction error accounts of learning are typically framed in terms of the immediate context: They assume that listeners generate predictions from this context, and

then learn by comparing these predictions to subsequent input, so that error signals can feedback and immediately adjust future processing (e.g., Arnal & Giraud, 2012). This framing is consistent with work on prediction in language comprehension, which presumes that listeners rapidly generate a variety of linguistic predictions. Listeners then use these predictions immediately, to guide interpretation of subsequent speech (Christiansen & Chater, 2016).

We have followed this framing when investigating the role of form and meaning predictions in Experiment 1 and 2. In particular, we have presumed that if these predictions enhance perception and learning, then they will do so immediately. For example, predictions made using the question *What colors are pandas*? should immediately affect interpretation of the distorted answer *Black and white*. Under these conditions, we have found that predictions about meaning affect perception and learning, while predictions about form do not.
But form predictions may play an important role in perception and learning when we relax the assumption that predictions are generated and used immediately. This finding would be theoretically important because it would not only indicate that predictions need not be now-or-never, as postulated by some theories of language processing (Christiansen & Chater, 2016) and assumed by a predictive coding account (e.g., Sohoglu et al., 2015), but it would also suggest that perception and learning can be enhanced by predictions about form. Thus in Experiment 3, we further tested how high-level knowledge enhances perception and learning by assessing the degree to which this enhancement actually depends on a match between the distorted input and an in-the-moment prediction. In particular, we assessed whether questions enhanced processing of distorted answers when this question was presented either three or six trials previously.

A few considerations suggest that form predictions may affect perception and learning (above-and-beyond semantic prediction effects) when questions and answers are separated in

time. First, we may have failed to find evidence for form prediction in Experiments 1 and 2 because of ceiling effects. In particular, the effect of semantic prediction in both of those studies may have been so large that it maximally enhanced perception (Experiment 1) and learning (Experiment 2), so that any effects of form were drowned out by the effects of meaning. By separating questions and answers in time, the relevant provision of information is more limited, and so there is more opportunity for both sources to contribute to perception. Moreover, precise form predictions will be more stable in memory over time than semantic predictions, perhaps because there is less likely to be interference between one precisely predicted wordform and another than between predicted semantic spaces. For example, Meyer & Schriefers (1991) found that although semantically similar representations compete with one another, phonologically similar representations do not. As a result, listeners may find it easier to comprehend a distorted answer when they have predicted its form many trials previously, than when they predict only its semantic space. Finally, there is some evidence that listeners predict meaning faster than they predict form. For example, Ito, Corley, Pickering, Martin, and Nieuwland (2016) found that prediction of word forms occurred only when participants read sentences at a very slow rate (700 ms SOA vs. 500 ms SOA), while prediction of semantics occurred at both SOAs. Together, these considerations suggest that separating questions and answers in time (and thus separating high-level knowledge from the distorted input) will provide a strong test of whether form predictions can affect perception and learning above-and-beyond semantic predictions.

We adopted this design in Experiment 3, in which participants heard the semantically inconsistent question-answer sequences from Experiments 1 and 2. However, we designed the stimulus lists so that the distorted answer heard on the current trial was a consistent response to a question heard either three or six trials previously (see Figure 1c). As in our previous experiments, we manipulated the form constraint of the question heard on a

previous trial, so it was either constraining and predicted a particular answer form (e.g., if listeners heard *What colors are pandas?* On trial one, then they would hear *Black and white* on trial four or seven), or unconstraining and did not predict a particular answer form (e.g., *What colors should I paint the wall*?).

However, we designed the stimulus lists so that the distorted answer heard on the current trial was a consistent response to a question heard either three or six trials previously (see Figure 1c). As in our previous experiments, we manipulated the form constraint of the previously heard question, so it was either constraining and predicted a particular answer form (e.g., if listeners heard *What colors are pandas*? on trial one, then they would hear *Black and white* on trial four or seven), or unconstraining and did not predict a particular answer form (e.g., *What colors should I paint the wall*?). Thus, we could determine whether form predictions enhanced perception when questions and answers are separated in time. If this is the case, then we expect listeners to be better at comprehending distorted answers when corresponding clear question presented many trials previously is form constraining rather than form unconstraining.

Method

Participants

One hundred and twenty-eight further native English speakers (41 males; *M*age = 26.71) were recruited from Prolific Academic using the same procedure as Experiment 1, and were randomly assigned to one of four stimulus lists.

Materials and Procedure

Experiment 3 used only the semantically inconsistent question-answer pairs from Experiments 1 and 2, and questions were either constraining or unconstraining. Unlike

previous studies, we presented each participant with all 31 question-answer pairs, as described in the Methods section for Experiment 1, so that inconsistent answers (e.g., *What colors are pandas? Tom Hanks*) could be primed by the presentation of their corresponding question on a previous trial (e.g., *Who voices the character Woody in the movie Toy Story? Buckingham Palace*). We then varied the trial distance between the presentation of the question and the presentation of its corresponding answer, so that the answer occurred either three or six trials after its question (see Figure 1c). Thus, participants were assigned to one of four stimulus lists in a between-participants design (form constraining three-primed, form constraining six-primed, form unconstraining three-primed, form unconstraining six-primed).

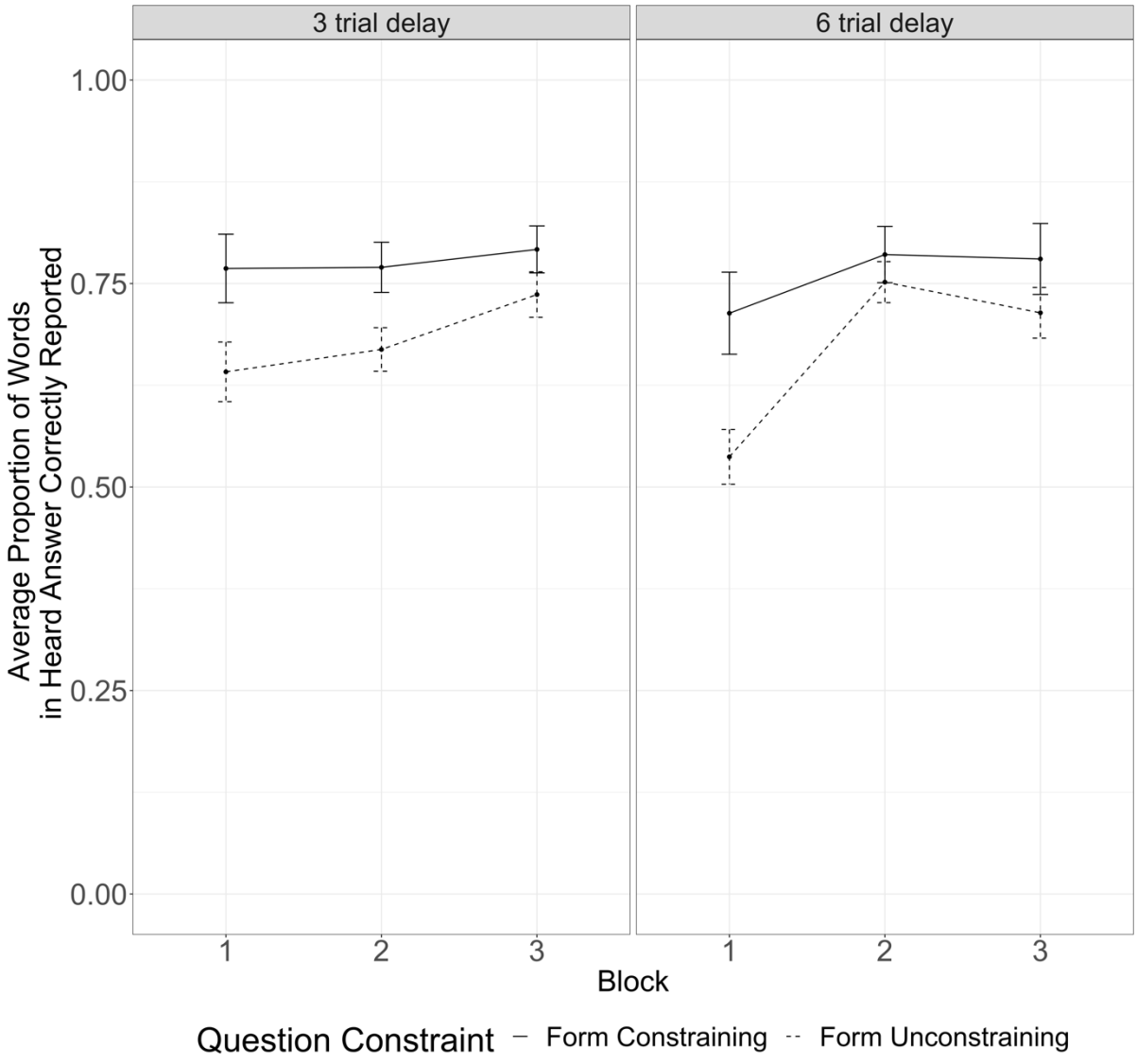
Note that not all answers could be primed in this design (e.g., the answer on the first trial could not be primed). This was true for six items in the three-priming list, and eight items in the six-priming list, and these items were excluded from data analysis.

Results and Discussion

As in Experiment 2, we analyzed participants' accuracy at reporting the heard distorted answers using a mixed effects logistic regression. Our model crossed three predictors: Question Constraint, Priming Distance (3 vs. 5; contrast coded as 1, -1), Block, and their full set of interactions. We also included by-participant random effects for Block and by-item random effects for Question Constraint, Priming Distance, and their interaction. All predictors were centered before being added to the model. Note that there were 31 trials in this experiment, in contrast with 15 in Experiments 1 and 2, and so Block was coded as a centered numeric predictor, with -1 for trials 1-9, 0 for trials 10-19, and 1 for trials 20-31. We discarded 23 trials (0.75%) because participants reported the question from the previous trial, rather than the answer for the current trial.

On average, participants correctly identified 73% (*SD* = 16%) of the words in the heard distorted answers (see Figure 6 for a breakdown of proportion by priming distance, question constraint, and block). Participants were better at identifying words in distorted answers in later than earlier blocks ($b = 0.52$, $SE = 0.13$, $p < .001$; see Table 5), suggesting that their ability to comprehend distorted speech increased with repeated exposure.

Figure 6. Observed means of the proportion of words in the heard answer reported correctly for the four factorial conditions across the three blocks in Experiment 3. Error bars represent ± 1 standard error from the mean.



We did not find any effect of Priming Distance: Participants were just as good at understanding distorted answers when they were primed by the presentation of their corresponding question six trials compared to three trials previously (b = 0.20, SE = 0.14, p = .15; see Table 5), suggesting that decay of predictions across trials is relatively limited. Thus, we found little evidence that in-the-moment predictions played a critical role in perception of noise-vocoded speech.

Table 5. Full model output for fixed effects for the analysis of word report scores in Experiment 3.

Predictor	Estimate (*SE*)	*z*	*p* value	Bayes factor
Intercept	1.71 (0.36)	4.77	< .001	-
Question Constraint	0.41 (0.16)	2.52	.01	3.50
Priming Distance	0.20 (0.14)	1.44	.15	0.35
Block	0.52 (0.13)	3.90	< .001	23.96
Question Constraint * Priming Condition	-0.11 (0.15)	-0.79	.44	0.19
Question Constraint * Block	-0.06 (0.15)	-0.54	0.59	0.01
Priming Condition * Block	-0.06 (0.12)	-0.51	0.61	0.14
Question Constraint * Priming Condition * Block	0.01 (0.12)	1.84	.40	0.14

In contrast to Experiments 1 and 2, we found that form predictions enhanced comprehension when distorted answers were presented many trials after their corresponding clear question. In particular, participants were better at reporting words in distorted answers

when their corresponding clear question was constraining and predicted a particular answer form, rather than unconstraining and only predicted the answer's semantic space (b = 0.41, SE = 0.16, p = .01). This effect of Question Constraint did not vary across our two Priming Distances (b = -0.11, SE = 0.15, p = .59), and the Bayes factor of 0.19 indicated strong evidence for this null effect. No further terms in the regression analysis were significant (see Table 5).

In addition, the plots of participants' average word report accuracy suggest that participants were better at comprehending semantically inconsistent distorted answers in this experiment (Experiment 3) than they were when reporting these inconsistent answers in Experiment 1 (Figure 7), perhaps because these semantically inconsistent answers received contextual support from previously presented questions. We assessed this effect statistically by fitting a mixed effects logistic regression, in which word report scores for the inconsistent answers were predicted by Question Constraint, Experiment (Priming vs. Perceptual Enhancement; contrast coded as 1, -1), Block, and their full set of interactions. Note that Block differs for the two experiments (15 trials in Experiment 1 and 31 in Experiment 3), and so we compared Experiment 1 to the first 15 trials of Experiment 3 so that Block was comparable across the two studies. We also included by-participant random effects for Block and by-item random effects for Question Constraint, Experiment, and their interaction. All predictors were centered before being added to the model.

As in our previous analysis of Experiment 3, participants were better at comprehending semantically inconsistent answers when they were preceded by form constraining rather than form unconstraining questions (b = 0.43, SE = 0.20, p = .03; see Figure 7 and Table 6). Participants were also better at comprehending semantically inconsistent distorted answers in Experiment 3 than in Experiment 1 (b = 1.67, SE = 0.34, p < .001), suggesting that hearing a matching question many trials before the presentation of its

corresponding distorted answer enhanced perception of that answer. We also found an interaction between Question Constraint and Experiment (b = 0.53, SE = 0.25, p = .01), such that participants were more accurate at reporting distorted answers when they had previously heard a constraining rather than an unconstraining question in Experiment 3 (b = 0.83, SE = 0.32, p = .009), but not in Experiment 1 (b = -0.28, SE = 0.32, p = .37). In other words, form predictions enhanced perception of distorted speech when there was a delay between question and answer (i.e., a delay between prediction and perception), but not when there was not.

Figure 7. Observed means of the proportion of words in the heard answer reported correctly for the semantically inconsistent answers across the first 15 trials in the priming experiment (Experiment 3) and all 15 trials in the perceptual enhancement experiment (Experiment 1). Error bars represent ± 1 standard error from the mean.

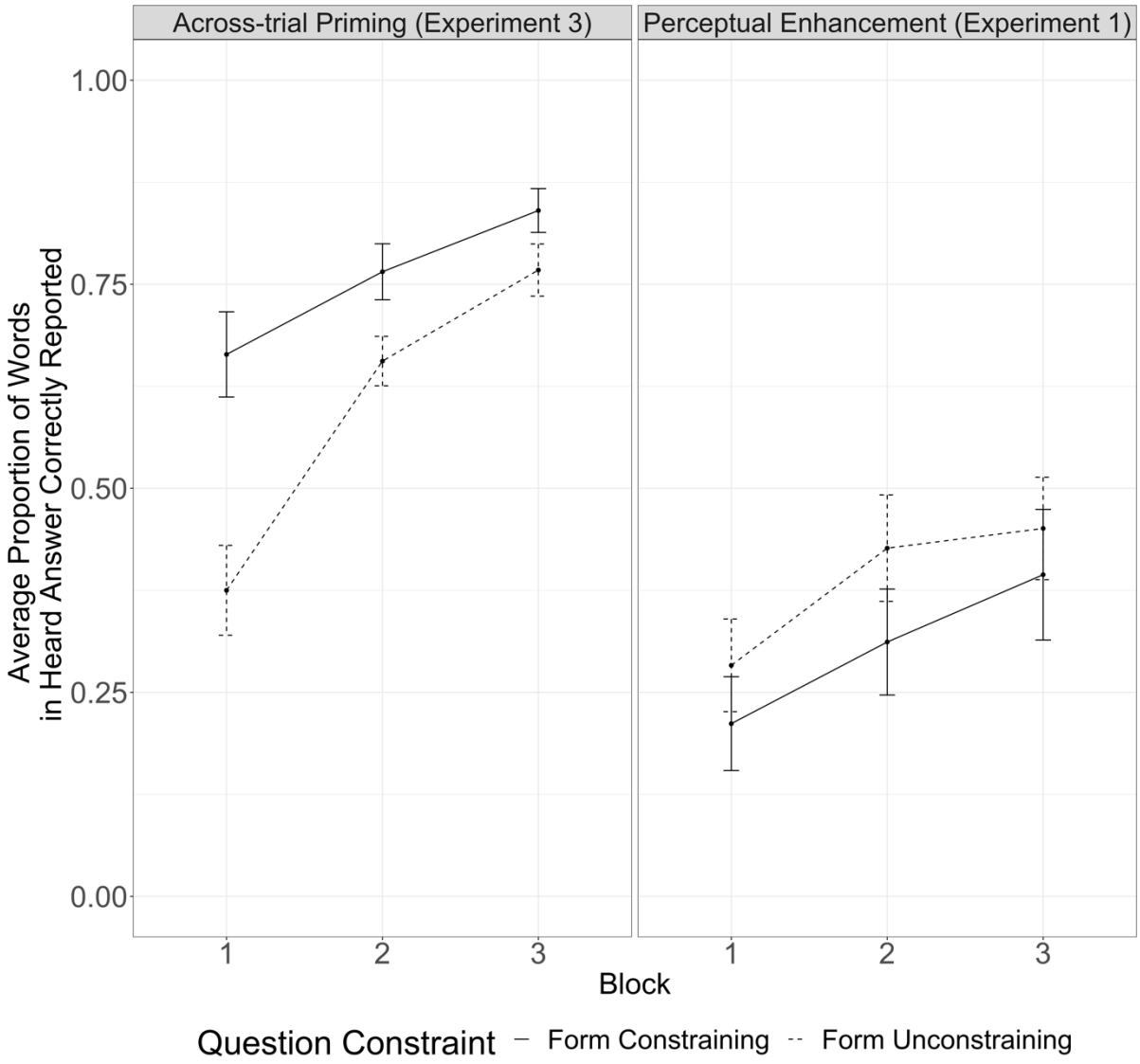


Table 6. Full model output for the fixed effects for the analysis comparing word report scores for semantically inconsistent answers in Experiments 1 and 3 (top subsection) and follow-up models testing the interaction between Question Constraint and Experiment (bottom subsection).

Full model output				
Predictor	Estimate (*SE*)	*z*	*p* value	Bayes factor
Intercept	0.73 (0.42)	1.74	.08	-
Question Constraint	0.43 (0.20)	2.17	.03	2.53
Experiment	1.67 (0.34)	4.90	< .001	>100
Block	0.72 (0.18)	3.98	<.001	>100
Question Constraint * Experiment	0.53 (0.21)	2.48	.01	5.62
Question Constraint * Block	-0.37 (0.12)	-2.97	.003	11.85
Experiment * Block	0.04 (0.16)	0.26	.80	0.58
Question Constraint * Experiment * Block	-0.32 (0.11)	-2.80	.005	6.77
Question Constraint * Experiment Interaction				
Perceptual Enhancement: Question Constraint	-0.28 (0.32)	-0.91	.37	0.40
Priming: Question Constraint	0.83 (0.32)	2.61	.009	8.81

General Discussion

Previous research demonstrates that high-level knowledge enhances perception of and learning about distorted speech. For example, listeners are better able to understand, and learn to understand, noise-vocoded sentences if they have previously heard or read a clear version of that sentence (e.g., Davis et al., 2005). In three experiments, we investigated how high-level knowledge enhances perception by presenting participants with question-answer sequences, in which the answer was noise-vocoded but the question was clearly spoken and could influence how the answer was processed. In particular, we varied the questions associated with each answer to test whether high-level knowledge enhances learning because listeners can use this knowledge to: (1) make highly specified target predictions of the form of the distorted words they are going to hear, or (2) to predict the high-level semantic space of the distorted input.

In Experiment 1, we found that participants were better at interpreting noise-vocoded answers when these answers were semantically consistent responses (e.g., *Black and* white) to a previously heard clear question (e.g., *What colors are pandas?*) than when they were semantically inconsistent responses (e.g., *Tom Hanks*), suggesting that being able to predict the likely semantic space of a distorted answer enhanced comprehension. Importantly, our signal detection analysis suggested that this effect occurred regardless of whether or not listeners could use the question to precisely predict the form of the distorted answer, thus suggesting that predictions about form did not enhance interpretation above-and-beyond predictions about meaning. Experiment 2 indicated that the same conclusions hold for perceptual learning. In particular, hearing vocoded answers that were semantically consistent responses to questions enhanced perceptual learning of vocoded speech, while predictions about form played no additional role in this process. These findings are inconsistent with predictive coding accounts of learning (e.g., Sohoglu et al., 2012), which claim that learning

to understand speech involves predicting the precise form of speech and then generating prediction errors (i.e., the match between the form of the predicted and actual input).

But in Experiment 3, we found that participants were better at understanding noise-vocoded answers when their corresponding clear question, presented either three or six trials previously, was form constraining rather than form unconstraining, suggesting that form predictions enhanced perception. In this study, perception was still enhanced even when questions were unconstraining (see Figure 6; accuracy was above 50%), suggesting that predictions about semantics still played an important role, even though listeners predicted form. This form effect contrasts with Experiments 1 and 2, which showed no evidence that form predictions affected interpretation and learning above-and-beyond the effects of semantics when distorted answers were presented immediately after their corresponding question.

The fact that both semantic and form-based predictions enhanced interpretation of distorted speech even over six trials is particularly interesting for predictive coding models. In particular, it indicates that when information is pre-activated, this pre-activation is long-lasting (lasting minutes at least). This finding contrasts with typical interpretations of predictive coding accounts, which assume that predictions are constantly updated to fit the current context (i.e., immediately predicting the next answer based on the just heard question), so that a prediction error can be immediately calculated based on the mismatch between predictions and the incoming linguistic input. The results of Experiment 3 are also inconsistent with other theories of prediction and processing (Christiansen & Chater, 2016), which suggest that expectations are generated and lost rapidly, to facilitate now-or-never language processing. By contrast, our data suggest that predictions (and even predictions about lower-level characteristics like form) can affect processing over long timescales.

Yet how should we account for the finding that more precise form predictions did not enhance perception or learning in Experiments 1 and 2, but did in Experiment 3? One possibility is that our manipulation of constraint was not strong enough in Experiments 1 and 2, and that this null effect is solely a task characteristic. For instance, perhaps participants' predictions about form were not precise enough to provide a facilitative boost. Alternatively, it could be the case that even unconstraining questions led participants to generate predictions about form that were precise enough to facilitate learning, and so the difference between constraining and unconstraining questions led participants to generate predictions about form that were precise enough to facilitate learning, and so the difference between constraining and unconstraining questions was too small to elicit a difference in accuracy. These explanations seem unlikely, however, because we did find effects of form constraint in Experiment 3.

It thus seems more likely that the discrepancy in the results of Experiments 1 and 2 and Experiment 3 do not reflect a confound in the stimuli used in this task, but rather says something important about the mechanisms that give rise to high-level influences on perception and learning. One possibility is that predictions about precise form may be less important for top-down processing of distorted speech than was previously suspected. In particular, form predictions may play no role in perception and learning, above and beyond the role of predictions about meaning, and so form prediction effects are measurable only when there is a smaller semantic prediction effect. In other words, the role of semantic predictions in perception (Experiment 1) and learning (Experiment 2) may have been so large that any effects of form were drowned out by effects of meaning. We observed form effects when questions and answers were separated in time (Experiment 3) because the relevant provision of information was more limited, and so there was more opportunity for both sources of information to contribute to processing.

One reason that form predictions had more of an effect in Experiment 3 than in Experiments 1 and 2 is that precise form predictions could be easier to maintain in memory over extended periods of time compared to semantic predictions. For example, predicted forms may not interfere with one another to the same degree that predicted semantic spaces do. Thus, separating questions and answers in time in Experiment 3 may have shifted the focus from semantic to form predictions, because these form predictions were easier to maintain. Another potential explanation for why form and semantic predictions had different levels of effectiveness over time is because listeners in Experiments 1 and 2 may not have had enough time to actually generate and implment form predictions. For example, there are suggestions that semantic predictions are faster to implement than form-based predictions (e.g., Ito et al., 2016). One implication of this is that the relatively rapid perceptual reorganization that was assessed in Experiments 1 and 2 (occurring over only 15 trials) may have been the result of fairly quick processes that only occurs at a high (e.g., semantic) level of representation, rather than at a low perceptual level. This would be consistent with evidence that there are two factors that affect variation in processing of distorted speech, one of which is lexico-semantic and one of which is acoustic (McGettigan, Rosen, & Scott, 2014).

That said, one potential concern about these conclusions is that participants' predictions may also be influenced by our experimental set up, such that participants in the semantically inconsistent answer conditions may have learned that the question was uninformative with regards to the semantics of the answer, and so may not have used these questions to generate predictions. However, participants in the inconsistent conditions in Experiment 1 still reported around 25% of the words in the answer that they expected to follow the question (i.e., the semantically consistent answer; see Figure 2, right panel), suggesting that listeners still used the questions to make some predictions, even though

questions explicitly misled participants. Future research could investigate whether listeners ignore high-level information if it is uninformative about the distorted input.

Nevertheless, our results offer a new perspective on previous research demonstrating that high-level knowledge, from clear auditory or written presentation of a stimulus prior to its distortion (e.g., Davis et al., 2005; Sohoglu et al., 2012) enhances perceptual learning. Where prior work demonstrated an important role for repetition, our results extend these findings by demonstrating that perception and learning can be enhanced solely through the activation of high-level semantic knowledge, and without requiring activation of lower-level form representations. The finding that even the activation of diffuse semantic features (from unconstraining questions) enhances processing and learning is consistent with previous speculation (Davis & Johnsrude, 2003) that the process of perceptual learning can be enhanced by any information source that can potentially constrain processing, whether low or high-level. A strong possibility is that this result indicates an important role for higher-level feedback, from semantics to the lexicon, in the learning process, rather than learning simply involving a comparison between predicted words and heard (distorted) words.

In sum, we have demonstrated that high-level knowledge of noise-vocoded speech facilitates perception and learning through the activation of high-level semantic features associated with the distorted input. We found little evidence that predictions about the form-features underlying that distorted input played an additional role under naturalistic conditions; these form predictions only supported comprehension and learning when there was a long delay and interference between the question and its corresponding distorted answer. These findings are inconsistent with accounts in which processing and learning of noise-vocoded speech are specifically enhanced by direct comparisons between the form of the predicted and actual input (i.e., predictive coding accounts), and instead suggest that

listeners use higher-level knowledge to predict the high-level semantic space of lower level input, which facilitates integration of, and learning about, novel distorted input.

Acknowledgements

Ruth Corps was supported by the Economic and Social Research Council [grant number ES/J500136/1]. Hugh Rabagliati was supported by grants from the Economic and Social Research Council [ES/L01064X/1] and the Leverhulme Trust [RPG-2014-253]. We thank Matthew Davis for sharing Matlab scripts used for vocoding.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in cognitive sciences*, *16*, 390-398.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*, 390-412.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). *Lme4: Linear mixed effects models using "Eigein" and S4* (R package version 1.1-14). Retrieved from http://CRAN.R-project.org/package=lme4.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*, 255-278.

Bent, T., Loebach, J. L., Phillips, L., & Pisoni, D. B. (2011). Perceptual adaptation to sinewave-vocoded speech across languages. *Journal of Experimental Psychology: Human perception and performance*, *37*, 1607-1616.

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS biology, 14,* http://dx.doi.org/10.1371/journal.pbio.1002577.

Bürkner, P-C. (2018). *Brms: Bayesian Regression Models using Stan* (R package version 2.1.0). Retrieved from https://CRAN.R-project.org/package=brms.

Caramazza, A. (1997). How many levels of processing are there in lexical access?. *Cognitive neuropsychology*, *14*, 177-208.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *31,* 489-558.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences, 36,* 181-204.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*, 3647-3658.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top–down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*, 3914-3932.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*, 3423-3431.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222-241.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological methods*, *3*, 186-205.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*, 391-407.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*, 1-12.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*, 283-321.

Doumas, L. A., & Martin, A. E. (2018). Learning structured representations from experience. *Psychology of Learning and Motivation*, *69*, 165-203.

Dorman, M. F., Hannley, M. T., Dankowski, K., Smith, L., & McCandless, G. (1989). Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant. *Ear and Hearing*, *10*, 44-49.

Dupoux, E. & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 914-927.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2*, 103-127.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 640-654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*, 13-42.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., et al. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cerebral Cortex, 14, 247–255.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 460-474.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157-171.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*, 773-795.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, http://dx.doi.org/10.3389/neuro.06.004.2008.

Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T. & Harris, H. D. (2018). Interaction in spoken word recognition models: Feedback helps. *Frontiers in Psychology, 9,* http://dx.doi.org/10.3389/fpsyg.2018.00369.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71-102.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McGettigan, C., Rosen, S., & Scott, S. K. (2014). Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: implications for cochlear implantation. *Frontiers in systems neuroscience*, *8*, http://dx.doi.org/10.3389/fnsys.2014.00018.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*, 1113-1126.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 1146-1160.

Miller, J. L. & Liberman, A. M. (1979). Some effects of later-occuring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457-465.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299-325.

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & cognition*, *26*, 844-851.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in cognitive sciences*, *11*, 105-110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, *36*, 329-347.

Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68,* 180-198.

Sebastián-Gallés, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & psychophysics*, *62*, 834-842.

Shannon, R. V., Fu, Q. J., & Galvin J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngologica*, *124*, 50-54.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303-304.

Signoret, C., Johnsrude, I., Classon, E., & Rudner, M. (2018). Combined effects of form-and meaning-based predictability on perceived clarity of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *44*, 277-285.

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, *113*, E1747-E1756.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*, 8443-8453.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, *30*, 415-433.

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage*, *60*, 1490-1502.

Wright, D. B., Horry, R., Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods, 41,* 257-267.

Appendix A: List of stimuli used in all three experiments. Note that Experiment 3 used only the semantically inconsistent conditions

Table A1: Stimuli used in Experiments 1-3, broken down by question constraint and answer plausibility.

Question Constraint	Question	Answer Consistency	Answer
Form Constraining	As well as cheese and tomato, which two toppings are usually on a Hawaiian pizza?	Semantically Consistent	Ham and pineapple
		Semantically Inconsistent	December twenty fifth
Form Unconstraining	What would you like for dinner?	Semantically Consistent	Ham and pineapple
		Semantically Inconsistent	December twenty fifth
Form Constraining	At which train station will you find platform nine and three quarters?	Semantically Consistent	Kings Cross

		Semantically Inconsistent	It hit an iceberg
Form Unconstraining	Where are you getting a train from?	Semantically Consistent	Kings Cross
		Semantically Inconsistent	It hit an iceberg
Form Constraining	How did The Titanic sink?	Semantically Consistent	It hit an iceberg
		Semantically Inconsistent	Andy Murray
Form Unconstraining	What happened to your boat?	Semantically Consistent	It hit an iceberg
		Semantically Inconsistent	Andy Murray
Form Constraining	How often does the dentist tell you to brush your teeth?	Semantically Consistent	Twice a day

		Semantically Inconsistent	Big Ben
Form Unconstraining	How often do you go outside for a walk?	Semantically Consistent	Twice a day
		Semantically Inconsistent	Big Ben
Form Constraining	What are the names of Ron Weasley's Mum and Dad?	Semantically Consistent	Molly and Arthur
		Semantically Inconsistent	The Tube
Form Unconstraining	What are your parents called?	Semantically Consistent	Molly and Arthur
		Semantically Inconsistent	The Tube
Form Constraining	What is Aurora Borealis commonly known as?	Semantically Consistent	The Northern Lights

		Semantically Inconsistent	Ham and pineapple
Form Unconstraining	What can you see out of your window?	Semantically Consistent	The Northern Lights
		Semantically Inconsistent	Ham and pineapple
Form Constraining	What colors are pandas?	Semantically Consistent	Black and white
		Semantically Inconsistent	Tom Hanks
Form Unconstraining	What colors should I paint the wall?	Semantically Consistent	Black and white
		Semantically Inconsistent	Tom Hanks
Form Constraining	What is London's underground railway also known as?	Semantically Consistent	The Tube

		Semantically Inconsistent	Hillary Clinton
Form Unconstraining	What is your least favorite method of transport?	Semantically Consistent	The Tube
		Semantically Inconsistent	Hillary Clinton
Form Constraining	What is the longest river in the world?	Semantically Consistent	The Amazon River
		Semantically Inconsistent	Snow White
Form Unconstraining	Where did you go swimming yesterday?	Semantically Consistent	The Amazon River
		Semantically Inconsistent	Snow White
Form Constraining	What is the name of the British prime minister?	Semantically Consistent	Theresa May

		Semantically Inconsistent	New York
Form Unconstraining	Who did you see when you visited London?	Semantically Consistent	Theresa May
		Semantically Inconsistent	New York
Form Constraining	What is the thirty first of December?	Semantically Consistent	New Year's Eve
		Semantically Inconsistent	Harry Potter
Form Unconstraining	When would you next like to go for drinks?	Semantically Consistent	New Year's Eve
		Semantically Inconsistent	Harry Potter
Form Constraining	When do you celebrate Christmas?	Semantically Consistent	December twenty fifth

		Semantically Inconsistent	A knife and fork
Form Unconstraining	When is your birthday?	Semantically Consistent	December twenty fifth
		Semantically Inconsistent	A knife and fork
Form Constraining	When do you celebrate Halloween?	Semantically Consistent	October thirty first
		Semantically Inconsistent	Robin Hood
Form Unconstraining	When do you next have a day off work?	Semantically Consistent	October thirty first
		Semantically Inconsistent	Robin Hood
Form Constraining	Where does Father Christmas live?	Semantically Consistent	The North Pole

		Semantically Inconsistent	New Year's Eve
Form Unconstraining	Where would you like to go on holiday?	Semantically Consistent	The North Pole
		Semantically Inconsistent	New Year's Eve
Form Constraining	Where does the president of America live?	Semantically Consistent	The White House
		Semantically Inconsistent	Molly and Arthur
Form Unconstraining	Where would you like to go when you visit America?	Semantically Consistent	The White House
		Semantically Inconsistent	Molly and Arthur
Form Constraining	Where does the prime minister live?	Semantically Consistent	Ten Downing Street

		Semantically Inconsistent	James Bond
Form Unconstraining	Where would you like to go today?	Semantically Consistent	Ten Downing Street
		Semantically Inconsistent	James Bond
Form Constraining	Where does the Queen live?	Semantically Consistent	Buckingham Palace
		Semantically Inconsistent	Black and white
Form Unconstraining	Which tourist attraction would you like to visit in London?	Semantically Consistent	Buckingham Palace
		Semantically Inconsistent	Black and white
Form Constraining	Which character starred in the famous 007 films?	Semantically Consistent	James Bond

		Semantically Inconsistent	Kings Cross
Form Unconstraining	What is your favorite film?	Semantically Consistent	James Bond
		Semantically Inconsistent	Kings Cross
Form Constraining	Which city is the Statue of Liberty in?	Semantically Consistent	New York
		Semantically Inconsistent	Buzz Lightyear
Form Unconstraining	Where would you like to go shopping?	Semantically Consistent	New York
		Semantically Inconsistent	Buzz Lightyear
Form Constraining	Which cutlery should I use to cut my food?	Semantically Consistent	A knife and fork

		Semantically Inconsistent	Theresa May
Form Unconstraining	What did you buy from the shop?	Semantically Consistent	A knife and fork
		Semantically Inconsistent	Theresa May
Form Constraining	Which famous clock is in London?	Semantically Consistent	Big Ben
		Semantically Inconsistent	Twice a day
Form Unconstraining	What is your brother's nickname?	Semantically Consistent	Big Ben
		Semantically Inconsistent	Twice a day
Form Constraining	Which female candidate recently ran for president of the United States?	Semantically Consistent	Hillary Clinton

		Semantically Inconsistent	The Northern Lights
Form Unconstraining	Who did you see when you visited America?	Semantically Consistent	Hillary Clinton
		Semantically Inconsistent	The Northern Lights
Form Constraining	Which fictional character lived with seven dwarves?	Semantically Consistent	Snow White
		Semantically Inconsistent	The Thames
Form Unconstraining	Who is your favorite fictional character?	Semantically Consistent	Snow White
		Semantically Inconsistent	The Thames
Form Constraining	Which river runs through London?	Semantically Consistent	The Thames

		Semantically Inconsistent	Donald Trump
Form Unconstraining	Where did you go on your boat ride yesterday?	Semantically Consistent	The Thames
		Semantically Inconsistent	Donald Trump
Form Constraining	Which space ranger starred in Toy Story?	Semantically Consistent	Buzz Lightyear
		Semantically Inconsistent	The Eiffel Tower
Form Unconstraining	Who is your favorite animated character?	Semantically Consistent	Buzz Lightyear
		Semantically Inconsistent	The Eiffel Tower
Form Constraining	Which tall building is in Paris?	Semantically Consistent	The Eiffel Tower

		Semantically Inconsistent	October thirty first
Form Unconstraining	Where are you going at Christmas?	Semantically Consistent	October thirty first
		Semantically Inconsistent	The Eiffel Tower
Form Constraining	Which young wizard defeated Lord Voldemort?	Semantically Consistent	Harry Potter
		Semantically Inconsistent	The White House
Form Unconstraining	What is your favorite book?	Semantically Consistent	Harry Potter
		Semantically Inconsistent	The White House
Form Constraining	Who is the best Scottish tennis player?	Semantically Consistent	Andy Murray

		Semantically Inconsistent	The North Pole
Form Unconstraining	Who is your favorite sportsman?	Semantically Consistent	Andy Murray
		Semantically Inconsistent	The North Pole
Form Constraining	Who is the newly elected president of America?	Semantically Consistent	Donald Trump
		Semantically Inconsistent	The Amazon River
Form Unconstraining	Who did you see interviewed on television recently?	Semantically Consistent	Donald Trump
		Semantically Inconsistent	The Amazon River
Form Constraining	Who leads a gang of outlaws in Sherwood Forest?	Semantically Consistent	Robin Hood

		Semantically Inconsistent	Ten Downing Street
Form Unconstraining	What is your best friend called?	Semantically Consistent	Robin Hood
		Semantically Inconsistent	Ten Downing Street
Form Constraining	Who voices the character Woody in the movie Toy Story?	Semantically Consistent	Tom Hanks
		Semantically Inconsistent	Buckingham Palace
Form Unconstraining	Who is your favorite actor?	Semantically Consistent	Tom Hanks
		Semantically Inconsistent	Buckingham Palace